

Mr. DLib: Recommendations-as-a-Service (RaaS) for Academia

Joeran Beel
Trinity College Dublin
ADAPT Centre, Ireland
j@beel.org

Bela Gipp
University of Konstanz
Germany
bela.gipp@uni-konstanz.de

Akiko Aizawa
National Institute of Informatics
Tokyo, Japan
aizawa@nii.ac.jp

ABSTRACT

Recommender systems for research papers are offered only by few digital libraries and reference managers, although they could help users of digital libraries etc. to better deal with information overload. One reason might be that operators of digital libraries do not have the resources to develop and maintain a recommender system. In this paper, we introduce Mr. DLib's recommender-system as-a-service. Mr. DLib's service allows digital libraries and reference managers to easily integrate a recommender system. The effort is low, and no knowledge about recommender systems is required. Mr. DLib's first pilot partner is the digital library Sowiport. Between September 2016 and February 2017, Mr. DLib delivered 60 million recommendations to Sowiport with a click-through rate of 0.15% on average. Mr. DLib is open source, non-profit, and supports open data.

KEYWORDS

Recommender system, digital library, recommendation-as-a-service, RaaS, API, web service, content-based filtering

1 INTRODUCTION

Recommender systems in academia help scientists to deal with information overload. However, only few operators of academic products – such as digital libraries and reference managers – offer recommender systems to their users. One potential reason why operators of academic products do not offer recommender systems, are the high costs for developing and maintaining them. Another reason might be that many operators do not have the skills to build a recommender system.

In this paper, we introduce Mr. DLib's recommender-system as-a-service that allows operators of academic products to easily integrate a recommender system in their products. The development and maintenance effort is minimal for an operator, and no knowledge about recommender systems is required.

There are a few other providers who offer recommendations as-a-service in Academia. BibTip [11] and bX [6] are commercial products. They use co-occurrence based recommendations, which is a generic approach not tailored to research papers, applicable only for a part of the documents in a library's catalogue, and suitable rather for larger libraries [3]. CORE [9, 10] and Babel [12] offer recommendations-as-a-service through an API, JavaScript client, and browser plug-in respectively. Babel is also considering

to develop add-ons for reference managers such as Zotero. Both services offer recommendations for open-access documents (CORE indexed around 68 million documents, Babel around 40 million). Babel also states they are welcoming other researchers to evaluate novel algorithms in Babel.

While each of the services has its strength and weaknesses, none offers all the following features: open-source, not-for-profit, open for research, integration in both digital libraries (of any size) and reference managers, recommendations for open access articles and the option for partners to have private document collections, and capable of recommending all documents in a corpus (100% coverage). Mr. DLib is offering these features.

2 MR. DLIB'S RECOMMENDER SYSTEM

Mr. DLib is a non-profit and open-source project, originally developed as Machine-reable Digital Library at the University of California, Berkeley and introduced at JCDL 2011 [2]. Today, Mr. DLib is run by researchers, among others, from the Trinity College Dublin, and the University of Konstanz.

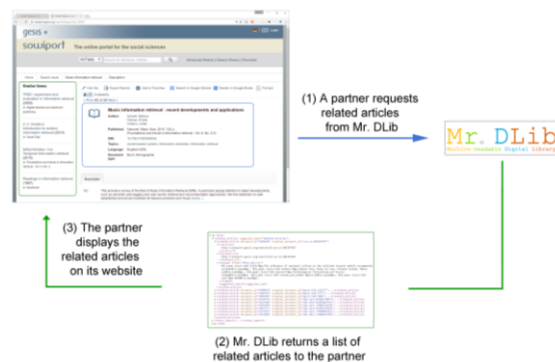


Figure 1: Illustration of the recommendation process

Mr. DLib's recommendations-as-a-service concept is illustrated in Figure 1. When a user browses the detail page of an article on a partner's web site, the partner requests a list of related articles from Mr. DLib RESTful Web Service. The request is sent as HTTP GET request.

```
GET /v1/documents/{document_id}/related_documents/ 1
```

¹ Example: https://api-mr-dlib.org/v1/documents/gesis-bib-136994/related_documents/.

When receiving a request, Mr. DLib calculates a list of related articles, and returns the list as XML. The partner converts the XML to HTML and displays the recommendations on its website (or mobile app, or desktop application).² Before this process can start, Mr. DLib needs access to the metadata of the partner's documents.

Mr. DLib uses primarily Apache Lucene/Solr's *More-Like-This* function for calculating document relatedness. We are also experimenting with alternative recommendation approaches such as stereotype and most-popular recommendations [5]. We further experiment with machine translations to enable multi-lingual content-based filtering, key-phrase extraction to enhance the effectiveness of content based filtering, and a re-ranking of recommendations based on readership data from Mendeley. Details on the algorithms and the architecture will be published in the future.

The production recommender system api.mr-dlib.org and development system api-dev.mr-dlib.org each run on a dedicated server (i7-6700k, 32GB RAM, SSD hard drives). The development system is also used for resource intensive tasks including document indexing, machine translations, keyphrase extraction, and calculation of bibliometrics. The uptime of the servers is constantly monitored³, and the average response time to deliver recommendations is 682ms. We further have a beta system api-beta.mr-dlib.org that runs on a virtual machine (4 cores, 14 GB RAM).

3 MR. DLIB'S PARTNER

Mr. DLib's first partner is the digital library Sowiport⁴, which is Germany's largest social science repository, operated by the GESIS institute [8]. Mr. DLib has indexed around 10 million documents from Sowiport. While GESIS agreed to have its documents recommended on other partners' websites, GESIS chose to recommend only its own content on Sowiport.

Since September 2016, Mr. DLib has delivered 60,836,800 recommendations to Sowiport, and users have clicked 91,545 of them. This equals on overall click-through rate (CTR) of 0.15%. Figure 2 shows the number of delivered recommendations and CTR by month. CTR is rather low and there is a notable variance among the months (e.g. 0.21% in September and 0.10% in December). The variance may be caused by different algorithms we are using. In addition, recommendations are delivered when web spiders such as the Google Bot crawl the Sowiport website. In contrast, clicks are logged with JavaScript, which is usually not executed by web spiders. Consequently, the CTR for 'human interaction' would be higher.

4 LICENSE AND POLICY

Mr. DLib advocates an 'open culture' and publishes source code as open source on GitHub⁵, describes project details in a public WIKI (Confluence)⁶, and manages issues in a public Ticket

Tracker (JIRA)⁷. Data relating to our research will be published on Harvard's Dataverse⁸ (as long as we can ensure privacy and copyrights of our partners). We also invite other researchers to evaluate their novel recommendation algorithms with Mr. DLib and our partners⁹.

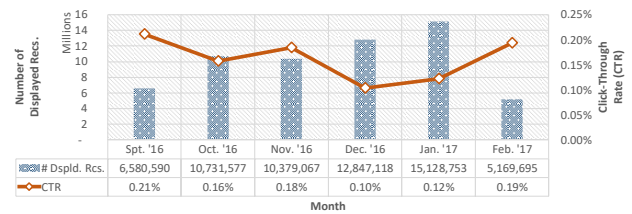


Figure 2: Number of Displayed Recommendations and CTR by Month (2016-09-08 to 2017-02-11)

5 OUTLOOK

Mr. DLib's recommender system was launched half a year ago and delivered more than 60 million recommendations in Sowiport. In the future, we will connect more partners (e.g. JabRef [7] and Docear [1, 4]), import more documents (e.g. the CORE corpus [10]), and improve the recommendation quality of Mr. DLib. We will also release a JavaScript client that allows an easier integration of Mr. DLib's recommender system into partner websites, and monitors the number of delivered and clicked recommendations more reliably. In the long run, we will extend the scope of Mr. DLib to not only recommend related articles, but personalized recommendations and items such as call for papers, research grants, and potential collaborators. In addition, we plan to establish a "living lab" with Mr. DLib that allows other researchers to evaluate their novel recommendation approaches in Mr. DLib's recommender system.

5 ACKNOWLEDGEMENTS

This work was supported by a fellowship within the FITweltweit programme of the German Academic Exchange Service (DAAD). This publication also has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under Grant Number 13/RC/2106. We are further grateful for the support provided by Zeljko Carevic, Philipp Mayr-Schlegel, Sophie Siebert, Siddharth Dinesh, Stefan Feyer, Dixon Daniel, Sara Mahmoud, Gabor Neusch, and Felix Beierle.

References

- [1] Beel, J. et al. 2011. Docear: An Academic Literature Suite for Searching, Organizing and Creating Academic Literature. *Proceedings of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries (JCDL)* (2011), 465–466.
- [2] Beel, J. et al. 2011. Introducing Mr. DLib, a Machine-readable Digital Library. *Proceedings of the 11th ACM/IEEE Joint Conference on Digital Libraries (JCDL '11)* (2011), 463–464.

² Example: <http://sowiport.gesis.org/search/id/gesis-bib-136994>

³ <http://monitoring.mr-dlib.org/>

⁴ <http://sowiport.gesis.org/>

⁵ <http://source-code.mr-dlib.org/>

⁶

⁶ <http://wiki.mr-dlib.org/>

⁷ <http://tickets.mr-dlib.org/>

⁸ <http://data.mr-dlib.org/>

⁹ <http://mr-dlib.org/research/>

- [3] Beel, J. et al. 2016. Research Paper Recommender Systems: A Literature Survey. *International Journal on Digital Libraries*. 4 (2016), 305–338.
- [4] Beel, J. et al. 2009. SciPlore MindMapping' - A Tool for Creating Mind Maps Combined with PDF and Reference Management. *D-Lib Magazine*. 15, 11 (Nov. 2009).
- [5] Beel, J. et al. 2017. Stereotype and Most-Popular Recommendations in the Digital Library Sowiport. *Proceedings of the 15th International Symposium of Information Science (ISI)* (2017).
- [6] ExLibris 2013. bX Usage-Based Services transform your discovery experience! *Web page*. <http://www.exlibrisgroup.com/category/bXUsageBasedServices>.
- [7] Feyer, S. et al. 2017. Integration of the Scientific Recommender System Mr. DLib into the Reference Manager JabRef. *Proceedings of the 39th European Conference on Information Retrieval (ECIR)* (2017).
- [8] Hienert, D. et al. 2015. Digital Library Research in Action – Supporting Information Retrieval in Sowiport. *D-Lib Magazine*. 21, 3/4 (2015).
- [9] Knoth, P. et al. 2017. Towards effective research recommender systems for repositories. *Proceedings of the Open Repositories Conference* (2017).
- [10] Knoth, P. and Zdrahal, Z. 2012. CORE: three access levels to underpin open access. *D-Lib Magazine*. 18, 11/12 (2012).
- [11] Monnich, M. and Spiering, M. 2008. Einsatz von BibTip als Recommendersystem im Bibliothekskatalog. *Bibliotheksdienst*. 42, 1 (2008), 54–54.
- [12] Wesley-Smith, I. and West, J.D. 2016. Babel: A Platform for Facilitating Research in Scholarly Article Discovery. *Proceedings of the 25th International Conference Companion on World Wide Web* (2016), 389–394.